TWO COMPLEMENTARY REPRESENTATIONS OF MULTIPLE TIME SERIES IN STATE SPACE INNOVATION FORMS

by

Masanao Aoki*

ABSTRACT

This paper uses two alternative innovation representations (forward and backward) of state space model of multivariate weakly stationary time series to suggest estimators of system models and innovation noise covariances which are alternative to the estimators based on stochastic realizations. A scheme for iterative improvements of the initial estimates is also suggested.

UCLA Department of Economics
Working Paper #628
September 1991

*2263 Bunche Hall, Department of Economics
University of California, Los Angeles 90024-1477
e-mail aoki@econ.sscnet.ucla.edu

I would like to thank A. Havenner and J. Dorfman for a number of helpful discussions and comments.

1. Introduction

Needs for forecasting multivariate time series arise in many undertakings both in experimental and non-experimental disciplines. In non-experimental forecasting problems such as many problems in economics, we cannot affect the data we observe by choosing "input" variables. We must deal directly with data. Many algorithms developed in the engineering literature presuppose the input-output pairs are available, and are not therefore applicable in modelling non-experimental data.

A computationally efficient and stable procedure applicable in such situations has been proposed in Aoki (1983, 1987). The algorithm does not require data in input-output pairs, uses only the first and second moment information of the data, and is therefore well suited to process data generated in any non-experimental undertaking. The basic algorithm requires that underlying data generating processes are weakly stationary with rational spectral density matrices.[1]

A number of successful applications of the algorithms have been made since the initial announcement of the algorithm.[2] Since 1987 several clarifications and improvements have been made to the basic algorithm. Its relations with other estimating schemes such as the least squares method, instrumental variables, and the canonical correlation method have also been pointed out. Some of these later developments and clarifications are included in Aoki (1990).

The purpose of this paper is to present further developments and refinements made to the algorithm and to discuss a way for iteratively

---

[1] The algorithm can be extended to time varying data generating processes so long as they are asymptotically stable.

[2] See Cerchi and Havenner (1988), Criddle and Havenner (1988), and Dorfman and Havenner (1991).

improving the estimates of the state space innovation representations of the data generating processes. More specifically, we give a more detailed account than in Aoki (1991) on how the IV estimators for the state space models are related to the forward and backward innovations in state space form, show how the "nestedness" property enjoyed by the original algorithm can be restored to the models constructed by the IV estimators, and relate the iterative improvements of the estimates proposed in this paper to the EM algorithms as applied to state space models by Shumway and Stoffer (1982).

2. Data Generating Processes

We assume that the data sequence $\{y_t\}$, $t = 0,1,2,\ldots$ is a mean zero weakly stationary linear process of dimension p, having a full rank rational spectral density matrix. These assumptions mean that we can posit a dynamic system for the data in the form

$$x_{t+1} = Ax_t + u_t, \quad t = 0,1,2,\ldots \tag{1}$$
$$y_t = Cx_t + v_t$$

with an unobserved state vector x_t of a finite dimension n, and an asymptotically stable constant matrix A where u_t and v_t are mean zero, serially and mutually uncorrelated noise processes. The matrix C is $p \times n$ and is called the observation matrix. Since x_t is not observable we estimate it and transform (1) into a state space innovation form before we can estimate matrices A and C.[3] We must estimate n as well. For simpler exposition we treat it fixed and known. See Aoki (1989) or Aoki and Havenner (1991) on this point.

[3] Shumway and Stoffer (1982) applies the EM algorithm to (1).

3. Forward Innovation Representation

There are two ways to estimate x_t in (1). In this section we estimate it by its orthogonal projection onto the subspace spanned by past data y_s, $s \le t-1$. We use the notation

$$y_{t-1}^{-} = \begin{bmatrix} y_{t-1} \\ y_{t-2}^{-} \end{bmatrix}$$

which defined y_{t-1}^{-} recursively as the stacked vector y_{t-1}, over, $y_{t-2}, \ldots$. For the moment we regard it as an infinite-dimensional going back to data at remote past, in order not to worry about the initial (or terminal) conditions. We use the notation

$$z_t = \mathcal{E}(x_t | y_{t-1}^{-}) \tag{2}$$

to denote this orthogonal projection.

The next section introduces a "dual" representation in which x_t is projected on the subspace spanned by y_τ, $\tau \ge t+1$.[4]

Denote the covariance matrix of this stacked vector by

$$R_{-} = \text{cov } y_{t-1}^{-}.$$

This matrix is also of infinite-dimensional, although in any actual implementation, we cut off the stacked vector at y_{t-K} for some positive integer K. Then R_{-} is $Kp \times Kp$.

Define a matrix Ω by

$$\Omega = \text{cov}(x_t, y_{t-1}^{-}).$$

These matrices are constant by the assumed weak stationarity of the data

---

[4]In a weakly stationary process the data covariance matrices depend on the time-differences only. There is no intrinsic direction of time for these statistics.

series. Then the state vector (2) is given by

$$z_t = \Omega R^{-1} y_{t-1}^{-}, \tag{3}$$

where the inverse is assumed to exist. At the moment (3) is not operational since we do not know the matrix Ω, which will be shown later to be determined by a suitable factorization of the Hankel matrix defined later. Advance the time index by 1 in (2) and note that the orthogonal projection has the property

$$\begin{aligned} \hat{E}(x_{t+1}|y_t^{-}) &= \hat{E}(x_{t+1}|y_{t-1}^{-}, y_t) \\ &= \hat{E}(x_{t+1}|y_{t-1}^{-}, e_t) \\ &= \hat{E}(x_{t+1}|y_{t-1}^{-}) + Be_t \end{aligned} \tag{4}$$

where

$$e_t = y_t - \hat{E}(y_t|y_{t-1}^{-})$$

is called the forward innovation component in the data vector y_t. Note that it is uncorrelated with y_{t-1}^{-} by construction, and that (e_t) are serially uncorrelated. Matrix B is defined by

$$B = E(x_{t+1}e_t')\Delta^{-1} \tag{5}$$

with

$$\Delta = \text{cov } e_t, \tag{6}$$

assumed to be positive definite. Since y_t is weakly stationary, the covariance matrix of the innovation vector is constant. The matrix B is also constant as we show later. The first term in (4) becomes, using the first of (1) and (2)

$$\begin{aligned} \hat{E}(Ax_t+u_t|y_{t-1}^{-}) &= A\hat{E}(x_t|y_{t-1}^{-}) \\ &= Az_t, \end{aligned}$$

where $\hat{E}(u_t|y_{t-1}^{-}) = 0$ by the assumption on the noise sequence. The set of equations in (1) is now replaced by

$$z_{t+1} = Az_t + Be_t \tag{7}$$

$$y_t = Cz_t + e_t$$

since $\mathcal{E}(v_t | y_{t-1}^-) = 0$, again by assumption. In (7), we assume that the dimension of the state vector is such that the pair (A,C) is observable.[5]

In this representation it is important to recognize that the state vector z_t is uncorrelated with the innovation vector e_t. Such a representation is called a forward innovation representation of the data generating process (1). Note that $\{z_t\}$ is also mean zero and weakly stationary. We denote its covariance matrix by Π. Then from (7) it satisfies the relation

$$\Pi = A\Pi A' + B\Delta B' \tag{8}$$

with $\Lambda_0 = C\Pi C' + \Delta$, where $\Lambda_0 = \text{cov } y_t$. When Δ is substituted out we see that it is a nonlinear equation in Π. It is called the algebraic Riccati equation.

Now return to the definition of the matrix Ω above (3). Since $\text{cov}(x_t, y_{t-1}^-) = \text{cov}(z_t, y_{t-1}^-)$, it is also a constant matrix

$$\Omega = [M \ A\Omega] = [M \ AM \ A^2M \ \ldots],$$

where we define the cross covariance matrix between z_t and y_{t-1} as

$$M = E(z_t y_{t-1}'). \tag{9}$$

To see that matrix Ω has the indicated structure observe that,

$$E(z_t y_{t-2}') = E(z_{t+1} y_{t-1}')$$

[5] Pair (A,C) is called observable if $Cz = 0$, $Az = \lambda z$ implies $z = 0$.

$$= AEz_t y'_{t-1}$$

$$= AM,$$

and so on. Matrix B defined by (5) can be rewritten as $B = E(z_{t+1}e'_t)\Delta^{-1}$ since e_t and e_{t+1} are uncorrelated by construction. Using (7) and (9), note that $M = A\Pi C' + B\Delta$. Thus matrix M appears only as an intermediate expression in defining matrix B, which is explicitly present in the time series dynamics (7). Section 4 provides a more direct interpretation of matrix M, however.

The (log) likelihood function can be written using (7), and maximized either directly to obtain the maximum likelihood estimates, or iteratively as in the EM algorithm. We follow an alternative route by noting that a factorization of the covariance matrix between y^-_{t-1} and y^+_t, defined as the stacked vector of all future observations $y_t, y_{t+1}, \ldots$ in that order, can be used to evaluate Ω and hence z_t as follows: Let

$$y^+_t = \begin{bmatrix} y_t \\ y^+_{t+1} \end{bmatrix}$$

Then from (7)

$$y^+_t = Oz_t + Se^+_t \qquad (10)$$

where the matrix O is the observability matrix of the model (7), i.e.,

$$O = \begin{bmatrix} C \\ CA \\ CA^2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

and S is a matrix of no immediate concern. Then the covariance matrix between the future and the past is

$$\text{cov}(y^+_t, y^-_{t-1}) = O\ \text{cov}(z_t, y^-_{t-1})$$

$$= \mathcal{O}\Omega.$$

The lefthand side defines a Hankel matrix made up of covariance matrices $\mathrm{cov}(y_{t+i}, y'_{t-j}) = \Lambda_{i+j}$ as submatrices $i = 0,1,\ldots,$ $j = 1,2,\ldots$ which is shown to be factorized as the product of two matrices $\mathcal{O}$ and Ω. The rank of matrix $\mathcal{O}$ is n which is the dimension of the state vector by the assumed observability of (7).

The sample covariance matrices

$$\hat{\Lambda}_\ell = T^{-1} \sum_{t=0}^{T-\ell} y_{t+\ell} y'_t, \quad \ell = 1,2,\ldots$$

are used to construct a sample Hankel matrix $\hat{H}$ which is factorized into $\hat{\mathcal{O}}$ and $\hat{\Omega}$ with ^ indicating sample versions. A numerically stable way for this factorization is the singular value decomposition

$$\hat{H} = U\Sigma V',$$

where

$$U'U = I_n,$$

$$V'V = I_n,$$

and

$$\Sigma = \mathrm{diag}(\sigma_1,\ldots,\sigma_n), \quad \alpha_1 \geq \sigma_2 \geq \ldots \geq \sigma_n > 0,$$

for some positive integer n, and we construct the two factors by

$$\hat{\mathcal{O}} = U\Sigma^{\frac{1}{2}} \tag{11}$$

and

$$\hat{\Omega} = \Sigma^{\frac{1}{2}} V'.$$

In the above the singular values smaller than the cutoff value σ_n are all replaced by zero and only the column vectors of U and V corresponding to the retained singular values are retained as U and V.

Other factorizations are possible such as $\hat{O} = U$ and $\Omega = \Sigma V'$ or $\hat{O} = U\Sigma^{\frac{1}{2}}$ and $\Omega = \Sigma^{\frac{1}{2}} V'$. These alternative forms amount to alternative choices of the coordinate system (or parameterization) of the state space. The representation that corresponds to the choice of (11) is called "balanced" since the gramians are equal

$$\hat{O}'\hat{O} = \hat{\Omega}\hat{\Omega}' = \Sigma.$$

Once matrix $\hat{\Omega}$ is obtained, we have estimates of the state vector z_t by (3) as

$$\hat{z}_t = \hat{\Omega}\hat{R}_-^{-1}y_{t-1}^-$$

where $\hat{R}_-$ is the estimates of R_- constructed from the sample covariance matrices, with $\hat{\Lambda}_{-\ell} = \hat{\Lambda}'_{\ell}$.

Once matrix C in (7) is estimated, the estimate of the innovation vector $\hat{e}_t$ is obtained as $y_t - \hat{C}\hat{z}_t$. Since (e_t) is weakly stationary,

$$\min_C \Sigma_t(y_t-C\hat{z}_t)'\Sigma^{-1}(y_t-C\hat{z}_t)$$

yields

$$\hat{C}\Sigma_t\hat{z}_t\hat{z}_t' = \Sigma y_t\hat{z}_t'$$

which is the same as the least squares estimate and is also the IV estimate using $\hat{z}_t$ as instruments. From the definition (3), we estimate the covariance matrix cov z_t, i.e., solution of the Riccati equation (8) by

$$\hat{\Pi} = \hat{\Omega}\hat{R}_-^{-1}\hat{\Omega}'. \tag{12}$$

Then the estimate of $\hat{C}$ is obtained as

$$\begin{aligned}\hat{C} &= (\Sigma y_t\hat{z}_t')(\Sigma\hat{z}_t\hat{z}_t')^{-1} \\ &= \hat{\Gamma}\hat{R}_-^{-1}\hat{\Omega}'(\hat{\Omega}\hat{R}_-^{-1}\hat{\Omega}')^{-1}\end{aligned} \tag{13}$$

where

$$\hat{\Gamma} = T^{-1}\Sigma y_t y_{t-1}^{-\prime}$$
$$= [\hat{\Lambda}_1, \hat{\Lambda}_2, \hat{\Lambda}_3, \ldots] = \hat{H}_{1\bullet},$$

since this is the first row submatrix of the sample Hankel matrix $\hat{H}$. In Aoki (1987) matrix C is estimated by using y_{t-1}^{-} as instruments,[6] i.e., by

$$C\Sigma_t z_t y_{t-1}^{-\prime} = \Sigma_t y_t y_{t-1}^{-\prime}$$

which leads to

$$C\hat{\Omega} = \hat{\Gamma},$$

and is solved using the pseudo-inverse as

$$\hat{C} = H_{.1}\hat{\Omega}^{+} \tag{14}$$

We remark that the error covariance matrix of (13) is not greater than that of (14) for ARMA processes, but is equal to (14) in the case of AR processes, see Aoki (1990, p.77).

An Alternative Derivation

We mention an alternative derivation of (7) due to Faurre (1979) since we use a similar derivation in the next section. With the infinite dimensional vector y_{t-1}^{-}, the definition of R_- can be partitioned and yields the relation

$$R_-^{-1} = \begin{bmatrix} \Lambda_0 & C\Omega \\ \Omega' C' & R_- \end{bmatrix}^{-1} = \begin{bmatrix} R^{11} & -R^{12} \\ -R^{21} & R^{22} \end{bmatrix}$$

since

[6] Note that y_{t-1}^{-} is uncorrelated with e_t by construction. Bowden and Turkington (1984), and Soderstrom and Stoica (1983) are books on IV estimators written by econometricians and systems analysts, respectively.

$$E(y_{t-1}y_{t-2}^{-\prime}) = CE(z_{t-1}y_{t-2}^{-\prime})$$

$$= CE(z_t y_{t-2}^{-\prime})$$

$$= C\Omega,$$

where the inversion formula yields

$$R^{12} = R^{11}C\Omega R_-^{-1} = (R^{21})'$$

and

$$R^{22} = R_-^{-1} + R_-^{-1}\Omega' C' R^{11} C R_-^{-1}.$$

Note that

$$R^{11} = (\Lambda_0 - C\Pi C')^{-1} = \Delta^{-1} \quad \text{from (8).}$$

Similarly

$$\Omega = E(z_t y_{t-1}^{-\prime})$$

$$= E[z_t(y_{t-1}' \; y_{t-2}^{-\prime})]$$

$$= [M \; Ez_t y_{t-2}^{-\prime}]$$

$$= [M \; Ez_{t-1} y_{t-.}^{-\prime}]$$

$$= [M \; A\Omega].$$

Then advance t by 1 in (3) to obtain

$$z_{t+1} = \Omega R_-^{-1} y_t^-$$

$$= [M \; A\Omega] \begin{bmatrix} R^{11} & -R^{12} \\ -R^{21} & R^{22} \end{bmatrix} \begin{bmatrix} y_t \\ y_{t-1}^- \end{bmatrix}$$

$$= M(R^{11}y_t - R^{12}y_{t-1}^-) + A\Omega(-R^{21}y_t + R^{22}y_{t-1}^-)$$

where the theoretical counterpart of (12) is used.

Then noting that $R^{12}y_{t-1}^- = R^{11}Cz_t$ and that $\Omega R^{21}y_t = \Pi C' R^{11} y_t$ we obtain

$$z_{t+1} = Az_t + B(y-Cz_t)$$

with

$$B = (M-A\Pi C')\Delta^{-1}, \tag{15}$$

in agreement with (7). In this forward innovation representation, matrices C and Π are the one which are mostly naturally estimated. Then the innovation covariance Δ is estimate by (8). To estimate B from (15) we need estimates of M and A. For these we use the dual representation of the data generating dynamics.

4. Backward Innovation Representation

Matrix M appears in the observation equation of the backward innovation representation. To see this, define

$$s_t = \mathcal{E}(x_t|y_t^+)$$
$$= \theta R_+^{-1} y_t^+$$

where

$$R_+ = \text{cov } y_t^+$$

and

$$\theta = \text{cov}(x_t, y_t^+) = \Pi O'$$

where (10) is used. It is convenient to scale s_t by Π^{-1} assuming that R_+ is positive definite to define a new state vector by

$$\zeta_t = \Pi^{-1} s_t = O' R_+^{-1} y_t^+.$$

In this representation time runs backward. Therefore, we express ζ_{t-1} as functions of ζ_t as follows. First, replace t by $t-1$ in the above,

$$\zeta_{t-1} = O R_+^{-1} y_{t-1}^+$$

where

$$y_{t-1}^{+} = \begin{bmatrix} y_{t-1} \\ y_t^{+} \end{bmatrix}.$$

Since $\mathcal{O}' = [C', A'\mathcal{O}]$, we express the above in the partitioned form

$$\zeta_{t-1} = \mathcal{O}'(C' \; A'\mathcal{O}')R_{+}^{-1} \begin{bmatrix} y_{t-1} \\ y_t^{+} \end{bmatrix}$$

$$= A'\zeta_t + (C'-A'ZM)S^{11}(y_{t-1}-M'\zeta_t)$$

where

$$R_{+}^{-1} = \begin{bmatrix} \Lambda_0 & M'\mathcal{O}' \\ \mathcal{O}M & R_{+} \end{bmatrix}^{-1} = \begin{bmatrix} S^{11} & -S^{12} \\ -S^{21} & S^{22} \end{bmatrix},$$

since using (10) $E(y_{t+1}^{+}y_t') = E(y_t^{+}y_{t-1}') = \mathcal{O}E(z_t y_{t-1}') = \mathcal{O}M$, with

$$S^{11} = \Lambda_0 - M'ZM,$$

$$Z = \mathcal{O}R_{+}^{-1}\mathcal{O},$$

$$S^{12} = S^{11}M'\mathcal{O}'R_{+}^{-1},$$

and

$$S^{22} = R_{+}^{-1} + R_{+}^{-1}\mathcal{O}MS^{11}M'\mathcal{O}'R_{+}^{-1}.$$

Now introduce the backward innovation vector by

$$b_{t+1} = y_t - \mathcal{E}(y_t|y_{t+1}^{+})$$

$$= y_t - M'\zeta_{t+1}.$$

By construction, b_t is uncorrelated with ζ_t. Then the backward dynamics are defined by

$$\zeta_{t-1} = A'\zeta_t + Pb_t \tag{16}$$

$$y_t = M'\zeta_{t+1} + b_{t+1}$$

where

$$P = C' - A'ZM,$$

$$\text{cov } b_t = \Delta_b = \Lambda_0 - M'ZM,$$

and Z is the solution of another Riccati equation

$$Z = A'ZA + G\Delta_b G'.$$

In this backward innovation representation

$$E(\zeta_t y_t') = A'ZM + G\Delta_b$$

from the two equations in (16). Using (10) and the observation equation in (7)

$$E(\zeta_t y_t') = C'$$

i.e., $$C' = A'ZM + G\Delta_b$$

which corresponds to (15) of the forward representation.

The covariances of the two state vectors are related to the canonical correlation coefficients between the future y_t^+ and the past y_{t-1}^-. Following Aoki (1991, Sec. 8.1) use normalized data vector $R_t^{-\frac{1}{2}} y_t^+$ and $R_-^{-\frac{1}{2}} y_{t-1}^-$ to do the singular value decomposition of the normalized Hankel matrix as

$$R_+^{-\frac{1}{2}} H R_-^{-\frac{1}{2}} = P\Gamma Q'$$

$$P'P = I, \quad Q'Q = I$$

where the matrix Γ is the diagonal matrix which displays the canonical correlation coefficients.

Then, recalling that $H = O\Omega$

$$\text{tr}(Z\Pi) = \text{tr}(O'R_+^{-1} O\Omega R Q_-^{-1}\Omega')$$

$$= \text{tr}(O'R_+^{-\frac{1}{2}} P\Gamma Q R_-^{-\frac{1}{2}}\Omega')$$

$$= \mathrm{tr}(P\Gamma Q'R_-^{-\frac{1}{2}}\Omega' O R_+^{-\frac{1}{2}})$$

$$= \mathrm{tr}(P\Gamma Q' Q\Gamma P')$$

$$= \mathrm{tr}\Gamma^2 .$$

See Table for a summary of the alternative expressions.

Estimates of M

In this representation M' is estimated exactly as C is estimated in the forward representation. The first factor of the sample Hankel matrix given in (11) yields the estimate $\hat{\zeta}_t$ of the backward state vector which is used as instruments to estimate M' in complete analogy with the estimation matrix C:

$$T^{-1}\Sigma y_t \hat{\zeta}'_{t+1} = \hat{M}'(T^{-1}\Sigma\zeta_{t+1}\zeta'_{t+1})$$

or

$$\hat{H}'_{\bullet 1} R_+^{-1} O = \hat{M}'(O' R_+^{-1} O)$$

or

$$\hat{M} = (O'R_+^{-1}O)^{-1} O' R_+^{-1} \hat{H}_{\bullet 1} \tag{17}$$

where $\hat{H}_{\bullet 1}$ is the first submatrix column of the sample Hankel matrix $\hat{H}$. This is contrasted with the estimator in Aoki (1987) which can be interpreted as using y^+_{t+1} as instrument in the second equation of (16).

Estimate of Matrix A

Unlike matrices C and M which appear explicitly in only one innovation representation, matrix A appears in both representations. We drop ^ from z_t, ζ_t, e_t and b_t from now on. Two alternative estimates are obtained by using z_t or ζ_t as instruments on the first equation in (7) or in (16). Multiply (7) from the right by z'_t and summing over t, we obtain

$$<z_{t+1}, z_t> = A<z_t, z_t> + <Be_t, z_t>$$

where $< , >$ denotes times average, e.g., $<Be_t, z_t> = T^{-1} \Sigma_{t=1}^{T} Be_t z_t'$. Dropping the last term since e_t and z_t are uncorrelated theoretically, an estimate of $\hat{A}$ of A is defined by

$$<z_{t+1}, z_t> = \hat{A}<z_t, z_t>. \tag{18}$$

Here

$$<z_t, z_t> = \hat{\Pi} = \hat{\Omega}\hat{R}_-^{-1}\hat{\Omega}',$$

and

$$<z_{t+1}, z_t> = \Omega\hat{R}_-^{-1} <y_t^-, y_{t-1}^-> \hat{R}_-^{-1}\Omega'.$$

When the transpose of (16) is multiplied from the left by $\hat{\zeta}_t$ and averaged over time, we obtain the dual of (18)

$$<\zeta_t, \zeta_{t-1}> = <\zeta_t, \zeta_t>A + <\zeta_t, Pb_t>$$

from which we construct another estimate of A by

$$<\zeta_t \zeta_{t-1}> = \hat{Z}\hat{A} \tag{19}$$

since $<\zeta_t, \zeta_t> = \hat{Z} = \hat{O}'\hat{R}_+^{-1}\hat{O}$.

Here

$$<\zeta_t \zeta_{t-1}> = \hat{O}'\hat{R}_+^{-1} <y_t^+, y_{t-1}^+>\hat{R}_|^{-1}\hat{O}$$

A third estimate is obtained by advancing t by one in (10)

$$\begin{aligned} y_{t+1}^+ &= Oz_{t+1} + Se_{t+1}^+ \\ &= OAz_t + OBe_t + Se_{t+1}^+ \end{aligned}$$

and multiplying it from the right by $\hat{z}_t'$ to derive

$$\langle y_{t+1}^{+}, z_t \rangle = O A \hat{\Pi} + O\langle B_t^{e}, \hat{z}_t \rangle + S\langle e_{t+1}^{+}, z_t \rangle .$$

The last two terms are theoretically zero and we define $\hat{A}$ by multiplying the above by $O' R_{+}^{-1}$ from the right

$$O' R_{+}^{-1} \langle y_{t+1}^{+}, z_t \rangle = \hat{Z}\hat{A}\hat{\Pi} \tag{20}$$

where we see that the left-hand side of (19) is equal to

$$\begin{aligned} O' R_{+}^{-1} \langle y_{t+1}^{+}, z_t \rangle &= O' R_{t+1}^{+} \langle y_{t+1}^{+}, y_{t-1}^{-} \rangle R_{-}^{-1} \Omega' \\ &= O' R_{+}^{-1} \hat{H}^{A} R_{-}^{-1} \Omega' \end{aligned} \tag{21}$$

where

$$\hat{H}^{A} = \langle y_{t+1}^{+}, y_{t-1}^{-} \rangle$$

is the sample Hankel matrix $\hat{H}$ shifted up by one submatrix (p) rows.

5. Properties of Estimators

Asymptotic Error Covariance Matrices of Estimators

In this section, asymptotic error covariance matrices of the IV estimators introduced in the previous sections are derived. The basic framework is the forward innovation model together with the assumption that the dynamic matrix A is asymptotically stable and the model is minimal, i.e., the pair (A,B) is a reachable pair. To evaluate the asymptotic behavior of these alternative estimates we use results due to Anderson and Kunitomo (1989) which generalized an earlier contribution by Billingsley (1961). Let $\mathcal{F}_t$ be an increasing sequence of σ-fields generated by z_t^{-} and e_t^{-}. Assume that $E(e_t | \mathcal{F}_{t-1}) = 0$, and $E(e_t e_t' | \mathcal{F}_{t-1}) = \Delta_t$. In addition we assume that, as a goes to infinity

$$\sup_t E[e_t' e_t I(e_t' e_t > a) | \mathcal{F}_{t-1}] \to 0 \quad \text{in probability.}$$

Throughout we use a version of the central limit theorem for martingale differences that the "vec" of $T^{-\frac{1}{2}} \Sigma_t z_t e'_t$ converges in distribution to a normal distribution function with mean zero and finite variance, under suitable conditions. Assuming some moment conditions such a theorem was proved in Billingsley [1961]. See also Lai and Wei [1985]. A more general version is obtained by Anderson and Kunitomo [1989] which allows for time varying conditional covariance matrices for the martingale differences $\{e_t\}$. We follow them in stating the next

__Lemma 1__ (Anderson and Kunitomo): Under the set of assumptions stated above, $T^{-1} \Sigma e_t e'_t$ converges in probability to Δ, $T^{-1} \Sigma e_t z'_t$ converges to zero in probability and that $T^{-1} \Sigma z_t z'_t$ converges in probability to Π.

If we further assume that

$$T^{-1} \sum_{t-1=\max(r,s)}^{T} (\Delta_t \otimes e_{t-r} e'_{t-s}) \to \delta_{r,s} (\Delta \otimes \Delta), \quad \text{in probability,}$$

__Lemma 2__: The expression $T^{-\frac{1}{2}} e_t z'_{t-i}$ is a martingale difference array and $T^{-\frac{1}{2}} \Sigma \text{vec} e_t z'_{t-i}$ converges to as $N(0, \Pi \otimes \Delta)$ as T goes to infinity, $i = 1, 2, \ldots,$ where $\Delta = \text{cov}\ e_t$ and $\Pi = \text{var}\ z_t$. Similarly $T^{-\frac{1}{2}} \text{vec} \Sigma e_t y'_{t-1}$ converges to $N(0, \Lambda_0 \otimes \Delta)$, where $\Lambda_0 = \text{var}\ y_t$.

We also use the next

__Theorem__ (Anderson and Kunitomo): The expression $T^{-\frac{1}{2}} \text{vec}(\Sigma_{t=1}^{T} z_t e'_t B')$ converges in distribution to a normal one with mean zero and covariance matrix $B \Delta B \otimes \Pi$ as T goes to infinity.

We first examine the estimator defined in (17). Let $\delta A = \hat{A} - A$. Then from (17)

$$\delta A \hat{\Pi} = \langle Be_t, z_t \rangle$$

or $$(\Pi \otimes I)\mathrm{vec}(\delta A) = T^{-1}\Sigma z_t \otimes Be_t .$$

Apply theorem to derive

$$\mathrm{cov}(\mathrm{vec}(\delta A)) = T^{-1}\hat{\Pi}^{-1} \otimes B\hat{\Delta}B' + O_p(T^{-1}).$$

Analogously, we derive from (19)

$$\hat{Z}\delta A = \langle \zeta_t, Pb_t \rangle$$

or $$(I \otimes \hat{Z})\mathrm{vec}\ \delta A = T^{-1}\Sigma Pb_t \otimes \zeta_t$$

from which we obtain

$$\mathrm{cov}(\mathrm{vec}(\delta A)) = T^{-1}P\Delta_b P' \otimes \hat{Z}^{-1} + O_p(T^{-1}).$$

To derive the error covariance matrix of the third estimator given by (20), rewrite it as

$$\hat{Z}\delta A\hat{\Pi} = \hat{Z} \langle Be_t, z_t \rangle + \langle e^+_{t+1}, z_t \rangle,$$

or $$\delta A\hat{\Pi} = \langle Be_t, z_t \rangle + W \langle e^+_{t+1}, z_t \rangle$$

with $$W = \hat{Z}^{-1}O'R_+^{-1} P.$$

Noting that the two terms on the right hand side are uncorrelated, we obtain, using $\mathrm{cov}\ e^+_{t+1} = I \otimes \Delta$.

$$\mathrm{cov}(\mathrm{vec}\delta A) = T^{-1} (\hat{\Pi}^{-1} \otimes B\Delta B' + \hat{\Pi}^{-1} \otimes W(I\otimes\Delta)W') + o_p(T^{-1})$$

A dual relation to (19) is obtained by noting that

$$y^-_{t-1} = \Omega'\zeta_t + Qb^+_{t+1}$$

and write y^-_{t-2} as

$$y^-_{t-2} = \Omega'\zeta_{t-1} + Qb^+_t$$
$$= \Omega'A'\zeta_t + \Omega'Pb_t + Qb^+_t.$$

Asymptotically, the first two are comparable and the third appears worse.

Their small sample properties, however, remain to be examined.

We next turn to analyze errors in estimating matrices C and M.

Matrix C: From (13), and the observation equation the estimation error matrix $\delta C = \hat{C} - C$ is given by

$$\delta C\hat{\Pi} = T^{-1}\Sigma e_t z_t'.$$

Using its vectorized expression

$$T^{-1/2}(\hat{\Pi}\otimes I)\text{vec }\delta C = T^{-1/2}\Sigma z_t \otimes e_t$$

and Lemma 2, we deduce the asymptotic error covariance matrix of δC to be

$$T(\hat{\Pi}\otimes I)\text{cov}(\text{vec }\delta C)(\hat{\Pi}\otimes I) \rightarrow \Pi \otimes \Delta,$$

i.e., $$T \text{ cov}(\text{vec}\delta C) \rightarrow \Pi^{-1} \otimes \Delta, \tag{22}$$

since as T goes to infinity, $\hat{\Pi} \rightarrow \Pi$.

In the case of the stochastic realization estimator, given in (14) the expression for δC is $\delta\hat{C}\hat{\Omega} = T^{-1}\Sigma e_t y_{t-1}^{-\prime}$. Instead of (22), we arrive at the expression

$$T(\hat{\Omega}'\otimes I)\text{cov }(\text{vec}\delta C)(\hat{\Omega}\otimes I) \rightarrow R_{-} \otimes \Delta, \tag{23}$$

or by taking the pseudo-inverse of $\hat{\Omega}$ given in (11)

$$T \text{ cov}(\text{vec}\delta C) \rightarrow \Xi\otimes\Delta,$$

where $\Xi = \Sigma^{-1/2}V'R_{-}V\Sigma^{-1/2}$. When the data generating process is AR(n), then we know that $\Omega R_{-}\Omega' = \Sigma^{1/2}V'R - V\Sigma^{1/2}$ where V and Σ are as in (11) is exact expression for Π, Aoki (1991, Sec. 7.2), then we can state

Proposition: In VAR models in which the minimal dimension of the model is an integer multiple of that of the data vector, the asymptotic covariance matrices of the stochastic realization estimator and the IV estimator with the state vector as instruments are the same.

This important property of the stochastic realization estimator follows by comparing (22) and (23), and noting that the state vector covariance matrix is given by $\Omega R^{-1}\Omega'$ in such situations and that the matrix V in the singular value decomposition of the Hankel matrix into $U\Sigma V'$ is such that VV' is the identity matrix so that $V(V'R_{-}^{-1}V)^{-1}V'$ is equal to R_{-}. Only when the state vector covariance matrices are not equal to $\Omega R_{-}^{-1}\Omega$, the asymptotic covariance matrix given in (22) is smaller than that given in (23) in the partial ordering of the symmetric positive definite matrices.

<u>Matrix M</u>: From (17) proceeding analogously we obtain

$$\delta\hat{M} = \hat{Z}^{-1}O'R_{+}^{-1}ST^{-1}\Sigma e_{t+1}^{+}y_{t}'.$$

Vectorizing this, the asymptotic error covariance matrix as T goes to infinity is given by

$$T\mathrm{cov}(\mathrm{vec}\,\delta M) \rightarrow \Lambda_0 \otimes W(I\otimes\Delta)W',$$

where $$W = Z^{-1}O'R_{+}^{-1}S,$$

where the matrix S is a lower triangular Toeplitz matrix with the identity matrix I_p on the main diagonal and the impulse response matrices $H_1, H_2, \ldots$ on the succeeding subdiagonal lines.

<u>Nestedness of Estimators</u>

The estimators of matrices A, C and M proposed in Aoki (1987) have the strict nestedness property with respect to the state vector dimension which are very useful. Once the estimates are made for dimension n^* the estimate of these matrices for any dimension for which the model is of minimal realization, then the estimate of these matrices for any dimension less than n^* can be read off as the appropriate submatrices of these matrices. See also Aoki and Havenner (1991, Sec. 3.2). The estimates obtained by using the state vectors as instruments do not enjoy this very desirable property. The strict nestedness can be restored once $\{\hat{z}_t\}$ is obtained with some dimension n^*.[7] To see this

define

$$\hat{\Gamma}_\ell = T^{-1} \sum_{t=1}^{N-\ell} y_{t+\ell}\hat{z}_t', \qquad \ell = 1,2,\ldots$$

and construct a new sample Hankel matrix $\hat{H}_z$ with $\hat{\Gamma}_\ell$'s in lieu of $\hat{\Lambda}_\ell$'s. The theoretical values are

$$\begin{aligned}\Gamma_\ell &= E(y_{t+\ell}z_t') \\ &= CA^{\ell-1}\Pi, \qquad \ell = 1,2,\ldots.\end{aligned}$$

Therefore, H_z admits a factorization

$$H_z = O\Omega_z$$

where matrix O is as before and we define

$$\Omega_z = [\Pi \; A\Pi \; A^2\Pi \ldots].$$

The singular value decomposition of H_z,

$$H_z = U_z\Sigma_z V_z',$$

leads to a balanced representation with

$$O_z = U_z\Sigma_z^{1/2}$$

and

$$\Omega_z = \Sigma_z^{1/2}V_z'.$$

Then the first submatrix row of H_z is $(H_z)_{1\bullet} = C\Omega_z$ and the relation

---

[7] This point was originally made in Havenner and Aoki (1987).

$$\hat{C} = (\hat{H}_z)_{1\bullet}\hat{\Omega}_z^+ = (\hat{H}_z)_{1\bullet}\hat{V}_z\hat{\Sigma}_z^{-\frac{1}{2}}$$

show that the estimate $\hat{C}$ possesses the nestedness property for $n \leq n^*$. Likewise for $\hat{A}$ derived from H_z. The estimate of Π now becomes nested since

$$(H_z)_{\bullet 1} = O_z\Pi$$

or

$$\hat{\Pi} = \hat{\Sigma}_z^{-\frac{1}{2}}\hat{U}_z'(\hat{H}_z)_{\bullet 1}.$$

Exact And Approximate Solutions of the Riccati Equation

In practice the size of the stacked vectors are finite and Hankel matrices are finite size truncations of the theoretical infinite-dimensional ones. Suppose that y_t^+ is truncated and y_{t+J} and y_{t-1}^- at y_{t-K}. Then $E(y_t^+ y_{t-1}^{-\prime})$ is $Jp \times Kp$. Now the state vector z_t being a summary of information contained in a truncated data vector $y_{t-1},\ldots,y_{t-k}$ in general fails to capture all information needed to forecast the future. The only important exception is a class of data generating processes which are autoregressive. If y_t is AR(m), then K can be taken to be m without loss of any information, while ARMA or MA processes admit no such exact finite truncations.

The Riccati equation for AR(m) processes with finite m admits closed form solutions as shown in Aoki (1991, Sec. 7.2) and hence an estimate of matrix B is obtained, once matrices A, C, M and Δ are estimated.

Even though the expression $\Pi = \Omega R_-^{-1}\Omega'$ is not exact with finite truncation for non-AR processes, they serve as an initial approximate expression of the covariance matrix of state vectors which are quite good in some cases. Aoki and Dorfman (1991) reports on a small scale Monte Carlo study to support this claim.

The only nontrivial calculation step in the algorithm is the computation of the covariance matrix Π which is minimal among the solutions of the algebraic Ricatti equation. It can be obtained iteratively or by converting the Riccati equation into a symplectic matrix equation. Even though the general procedure is known in the control and estimation literature, the details of the matrix differ in our stochastic realization estimation problem as shown by the symplectic matrix derived in Aoki (1987, 1991).[8] The symplectic matrix in time series may possess spurious unit roots when matrix ψ (defined in Aoki (1990, p. 79) is singular. A remedy was suggested in Aoki (1987, p. 124) by increasing the estimate of Λ_0. Vaccaro and Vukina (1991) also suggest similar methods to remedy the nonpositive condition. It is better, however, to use the (approximate) expression $\Omega R_{-}\Omega'$ for Π as an initial estimate and iteratively improve it either as a part of EM algorithm or as described next.

6. Improving Estimates

Estimated models described above produce quite satisfactory out-of-sample forecasts in applications as shown in Cerchi and Havenner (1988), Criddle and Havenner (1989) and others. In practical implementation of the algorithm, the stacked data vector is necessarily truncated. Truncated data vector lose information except in AR data generating processes, where AR(p) require only data $y_{t-1}, \ldots, y_{t-p}$, model matrices estimated by the algorithm are only approximate rendition of the theoretically derived expressions in the above. They can be further improved, if desired, by a procedure analogous to the EM algorithm reported in Shumway and Stoffer (1982) since the expression for the (log) likelihood function is available.

---

[8] One major difference is that the noise covariance matrices and model parameters are assumed known in the estimation literature, while they must be estimated in model construction.

After matrices A, B, and C are estimated as described in the previous sections, (7) is used together with $\hat{z}_0 = 0$ to produce $\{\hat{z}_t\}$ and $\{\hat{e}_t\}$. (In particular $\hat{\Delta}$ should be close to $\langle e_t, e_t \rangle$.) For asymptotically stable A, the initial condition z_0 does not matter too much. If desired, however, a "backcasting" procedure can be invoked to estimate $\hat{z}_0$. In what follows, we set $\hat{z}_0 = 0$ for simpler exposition of the basic iterative schemes for modifying estimates of the matrices. Denote the initial sequences by $\{z_t^0\}$ and $\{e_t^0\}$ to distinguish them from $\{z_t^i\}$ and $\{e_t^i\}$ to be introduced below.

In the version of the EM algorithm due to Shumway and Stoffer (1982), the model represented by (1) is used to write the log-likelihood function as

$$L = -\frac{1}{2}\ \ell n|Q| - \frac{1}{2}\ \mathrm{tr}\ Q^{-1}\ \Sigma_t(x_{t+1}-Ax_t)\ (x_{t+1}-Ax_t)'$$

$$-\frac{1}{2}\ \ell n|R| - \frac{1}{2}\ \mathrm{tr}\ R^{-1}\ \Sigma_t(y_t-Cx_t)\ (y_t-Cx_t)'$$

t const.

where

$$\mathrm{cov}\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \mathrm{diag}(Q,R).$$

In the innovation formulation, the representation (7) rather than (1) is employed to write the joint probability density and the log-likelihood as:

$$P(y_t^-, z_t^-) = P(y_t, z_t | y_{t-1}^-, z_{t-1}^-)P(y_{t-1}^-, z_{t-1}^-)$$

with

$$P(y_t, z_t | y_{t-1}^-, z_{t-1}^-) = P(z_t | y_{t-1}^-, z_{t-1}^-)P(y_t | y_{t-1}^-, z_t^-)$$

$$= P(z_t | z_{t-1})P(y_t | z_t)$$

with

$$P(z_t | z_{t-1}) = \mathrm{const}\ \frac{1}{|B\Delta B'|^{1/2}}\ \exp - \frac{1}{2}\ (z_t - Az_{t-1})'(B\Delta B')^{-1}(z_t - Az_{t-1})$$

and

$$P(y_t|z_t) = \text{const} \frac{1}{|\Delta|^{1/2}} \exp - \frac{1}{2} (y_t - Cz_t)'\Delta^{-1}(y_t - Cz_t).$$

The log-likelihood function becomes

$$L_{state} = - \frac{1}{2} \ell n|B\Delta B'| - \frac{1}{2} \ell n|\Delta|$$

$$- \frac{1}{2} \operatorname{tr}(B\Delta B')^{-1} \Sigma_t(z_t - Az_{t-1})(z_t - Az_{t-1})'$$

$$- \frac{1}{2} \operatorname{tr} \Delta^{-1}\Sigma_t(y_t - Cz_t)(y_t - Cz_t)',$$

(assuming that $z_0 = 0$).

Maximizing the above with respect to C and A yields, respectively

$$\langle y_t, z_t\rangle = C\langle z_t, z_t\rangle \tag{24}$$

and

$$\langle z_t, z_{t-1}\rangle = A\langle z_{t-1}, z_{t-1}\rangle. \tag{25}$$

When the log-likelihood function is maximized with respect to B, we get

$$B\Delta B' = \langle z_t - Az_{t-1}, z_t - Az_{t-1}\rangle. \tag{26}$$

The maximization with respect to Δ yields

$$0 = -\Delta^{-1} + \Delta^{-1} \langle y_t - Cz_t, y_t - Cz_t\rangle\Delta^{-1} \tag{27}$$

$$-B'(B\Delta B')^{-1}B + B'(B\Delta B')^{-1} \langle z_t - Az_{t-1}, z_t - Az_{t-1}\rangle(B\Delta B')^{-1}B$$

Substitute (26) into (27) to simplify it as

$$\Delta = \langle y_t - Cz_t, y_t - Cz_t\rangle. \tag{28}$$

Having $\{z_t^i\}$, $\{e_t^i\}$, (24) yields $\hat{C}^{i+1}$ by

$$\hat{C}^{i+1} = \langle y_t, z_t^i\rangle\langle z_t^i, z_t^i\rangle^{-1}$$

and $$\hat{A}^{i+1} = \langle z_t^i, z_{t-1}^i\rangle\langle z_{t--1}^i, z_{t-1}^i\rangle^{-1}.$$

The latter yields

$$z_{t+1}^{i+1} = \hat{A}^{i+1} z_t^{i+1} + \hat{B}^{i+1} e_t^{i+1} \tag{29}$$

with

$$e_t^{i+1} = y_t - \hat{C}^{i+1} z_t^{i+1}, \quad t = 0,1,\ldots \tag{30}$$

and $\hat{B}^{i+1} = \langle z_{t+1}^{i} - \hat{A}^{i} z_t^{i}, e_t^{i} \rangle \langle e_t^{i}, e_t^{i} \rangle^{-1}$.

Then (28) produces

$$\hat{\Delta}^{i+1} = \langle e_t^{i+1}, e_t^{i+1} \rangle \tag{31}$$

7. The Two-Step Procedure and Iterations

When the largest eigenvalue in magnitude of the dynamic matrix A is still less than one but close to it, the time series may appear nonstationary to naked eye and modeling of such time series should be carried out in two steps; first to model the dynamic mode corresponding to the largest eigenvalue in magnitude and then the remainder, as carried out in Aoki (1990, Sec. 11.5). This procedure has been applied with success to data series with the largest eigenvalue near .98 or .99 and the next largest one .93 or so, for example.

We write the log likelihood function in a decomposed form to derive expressions for their iterative improvements.

The data series is represented as

$$y_t = C\tau_t + Hz_t + e_t$$

where τ_t is the state variable for the dynamic mode with the largest eigenvalue ρ. Here to be definite and simplify explanation, we assume that only one eigenvalue $\rho > 0$ is such that $\rho >$ all other eigenvalue of A in magnitude. The state variable τ_t is scalar, and it evolves with time according to

$$\tau_t = \rho\tau_{t-1} + \psi z_{t-1} + be_{t-1}$$

The remaining dynamic modes contained in the data series are modeled by z_t, evolving with time by

$$z_t = Fz_{t-1} + Ge_{t-1}.$$

In this procedure both τ_t and z_t are measurable functions of y^-_{t-1}, and orthogonal to e_t. They jointly form the state vector.

This representation[9] is useful since it shows that a linear combination of components of y_t, $\nu' y_t$, where $\nu' C = 0$, does not contain the dynamic mode corresponding to the eigenvalue ρ. This procedure for forming $\nu' y_t$ is analogous to the notion of cointegration which corresponds to the data series with unit root.

The relevant part of the log-likelihood function is

$$- \frac{1}{2} \ell n|\Delta| - \frac{1}{2} \operatorname{tr} \Delta^{-1} \Sigma_t (y_t - C\tau_t - Hz_t)(y_t - C\tau_t - Hz_t)'$$

$$- \frac{1}{2} \ell n|b\Delta b'| - \frac{1}{2} \operatorname{tr}(b\Delta b')^{-1} \Sigma_t (\tau_t - \rho\tau_{t-1} - \psi z_{t-1})^2$$

$$- \frac{1}{2} \ell n|G\Delta G'| - \frac{1}{2} \operatorname{tr}(G\Delta G')^{-1} \Sigma_t (z_t - Fz_{t-1})(z_t - Fz_{t-1})'.$$

Its maximum with respect to C and H leads to

$$[\langle y_t, \tau_t \rangle \ \langle y_t, z_t \rangle] = [C,H] \operatorname{cov} \begin{pmatrix} \tau_t \\ z_t \end{pmatrix} \tag{32}$$

with

$$\operatorname{cov} \begin{pmatrix} \tau_t \\ z_t \end{pmatrix} = \begin{bmatrix} \langle \tau_t, \tau_t \rangle & \langle \tau_t, z_t \rangle \\ \langle z_t, \tau_t \rangle & \langle z_t, z_t \rangle \end{bmatrix}.$$

The maximization with respect to ρ and ψ leads to

---

[9]Its dual representation may be used instead, which is the error-correction form as shown in Aoki (1988). An EM-like algorithm can be developed as well.

$$[\langle \tau_t, \tau_{t-1} \rangle \ \langle \tau_t, z_{t-1} \rangle] = [\rho, \psi] \ \mathrm{cov} \begin{pmatrix} \tau_{t-1} \\ z_{t-1} \end{pmatrix}. \tag{33}$$

The dynamic matrix F is estimated by

$$\langle z_t, z_{t-1} \rangle = F \langle z_{t-1}, z_{t-1} \rangle. \tag{34}$$

The remaining parameters are estimated by

$$b\Delta b' = \langle \tau_t - \rho\tau_{t-1} - \psi z_{t-1}, \ \tau_t - \rho\tau_{t-1} - \psi z_{t-1} \rangle \tag{35}$$

$$G\Delta G' = \langle z_t - Fz_{t-1}, \ z_t - Fz_{t-1} \rangle \tag{36}$$

and

$$\Delta = \langle e_t, e_t \rangle \tag{37}$$

with

$$e_t = y_t - C\tau_t - Hz_t.$$

Estimates of b and G may be calculated by

$$b = \langle \tau_t - \rho\tau_{t-1} - \psi z_{t-1}, e_{t-1} \rangle \Delta^{-1} \tag{38}$$

and

$$G = \langle z_t - Fz_{t-1}, e_{t-1} \rangle \Delta^{-1}. \tag{39}$$

The iterative procedure is therefore entirely analogous to the one described in the previous subsection.

The two-step procedure is employed to generate the initial set of estimate of C, H, F, b, G and Δ, together with $\{e_t^0\}$, $\{\tau_t^0\}$ and $\{z_t^0\}$. Then (32) - (39) are used to update these estimates which in turn update the innovation and state vectors.

8. Discussions

This paper reviewed the state space modeling of vector-valued time series in Aoki (1987, 1990) from the perspectives of two alternative innovation representations and showed that each representation is natural for estimating matrices C and M respectively. The dynamic matrix A can be estimated in at least three ways.

Since approximate solutions to the Riccati equation produce corresponding approximate expressions of matrix B, and hence the approximate models, it is interesting to compare alternative forecasts implied by alternative approximate expressions for matrix B. We note that the closed form expression of matrix Π in (12) with the truncated matrices corresponding to a finite truncation of the stacked data vector y_{t-1}^{-} at y_{t-k} for some $K > 0$ produces one such matrix B.

TABLE

Model	$z_{t+1} = Az_t + Be_t$	$\zeta_{t-1} = A'\zeta_t + Pb_t$
	$y_t = Cz_t + e_t$	$y_t = M'\chi_{t+1} + b_{t+1}$
with		with
	$e_t = y_t - \mathcal{E}(y_t \mid y^-_{t-1})$	$b_t = y_t - \mathcal{E}(y_t \mid y^+_{t+1}$
State vector	$z_t = \Omega R_-^{-1} y^-_{t-1}$	$\zeta_t = \mathcal{O}' R_+^{-1} y^+_t$
	$\mathrm{cov}\ z_t = \Omega R_-^{-1} \Omega'$	$\mathrm{cov}\ \chi_t = \mathcal{O}' R_+^{-1} \mathcal{O}$
	$y^+_t = \mathcal{O} z_t + Se^+_t$	$y^-_{t-1} = \Omega' \zeta_t + Qb^+_t$

REFERENCES

Anderson, T.W., and N. Kumitomo, "Asymptotic Robustness in Regression and Autoregression Based on Lindberg Conditions," Stanford Univ. Tech. Rept. No. 23, June 1989.

Aoki, M., Note on Economic Time Series Analysis: System Theoretic Approach, Heidelberg: Springer-Verlag, 1983.

__________, State Space Modelling of Time Series, Heidelberg/NY: Springer-Verlag, 1987.

__________, State space Modelling of Time Series, 2nd revised ed., Heidelberg/NY: Springer-Verlag, enlarged 1990.

__________, "Singular Values, Canonical Correlation Coefficients and the Dimension of State Space Models of Time Series," to be presented at the 1989 NSF/Nber Seminar, Madrid, Sept. 1989.

__________, "Cointegration, Error Correction and Aggregation in Dynamic Models: A Comment," Oxford Bull. Econ. Stat., 50 (198): 89-95.

__________, "Instrumental Variables Estimators for State Space Models of Time Series," in G. Conte, A.M. Perdon, and B. Wyman (eds.), New Trends in System Theory, Boston: Birkhauser, 1991: 73-80.

__________ and A. Havenner, "State Space Modeling of Multiple Time Series," Econometric Reviews (1991).

__________ and J. Dorfman, "Statistical Analysis of Some Identification Procedures for State Space Innovation Models," presented at MTNS91 International Symposium on the Mathematical Theory of Networks and Systems, Kobe, Japan, forthcoming, 1991.

Billingsley, P., "The Lindberg-Lévy Theorem for Martingales," Proceedings of the American Mathematical Society, 12 (1961): 788-92.

Bowden, R.J., and D.A. Turkington, Instrumental Variables, Cambridge/NY: Cambridge University Press, 1984.

Cerchi, M., and A. Havenner, "Cointegration and Stock Prices," Journal of Economics Dynamic. Control, 12 (1988): 333-46.

Criddle, K., and A. Havenner, "Forecasting Halibut Biomass Using System Theoretic Time Series Methods," American J. Agri. Economics, 11 (1989): 422-31.

Dorfman, J.H., and A. Havenner, "State Space Modelling of Cyclical Supply, Seasonal Demand and Agricultural Inventories," American Agricultural Economics, 73 (1991): 839-40.

Faurre, P.L., M. Cleyet, and F. Germain, Operateurs Rationels Positifs, Paris: Durod, 1979.

Havenner, A., and M. Aoki, "Econometrics and Linear Systems Theory in Multivariate Time Series Analysis," Working Paper, UC-Davis Dept. of Agricultural Economics.

__________ and __________, "An Instrumental Variable Interpretation of Linear System Theory Estimation," J. Econ. Dyn. Control, 12 (1988).

Kraybill, D.S., and J.H. Dorfman, "A Dynamic Intersectoral Model of Regional Economic Growth," J. Regional Science, 32 (Feb. 1992), forthcoming.

Lai, T.L., and C.Z. Wei, "Asymptotic Properties of Multivariate Weighted Sums With Application to Stochastic Regression in Linear Dynamic Systems," in D.R. Krishnaiah (ed.), Multivariate Analysis, Amsterdam: North Holland, 1985.

Ljung, L., System Identification: Theory for the User, Englewood Cliffs, NJ: Prentice Hall, 1987.

Lindquist, A., and M. Pavon, "On the Structure of State Space Models for Discrete-Time Stochastic Vector Processes," IEEE AC-29 (1984): 418-32.

Sargan, D., Lectures on Advanced Econometric Theory, M. Desai (ed.), Oxford: Basil Blackwell, 1988.

Shumway, R.H., and D.S. Stoffer, "An Approach to Time Series Smoothing Forecasting Using the EM Algorithm," J. Time Series Anal., 3 (1982): 253-64.

Soderstrom, T., and P. Stoica, Instrumental Variable Methods for System Identification, Lecture Notes in Control and Information Sciences, New York: Springer-Verlag, 1983.

Vaccaro, R.J., and T. Vukins, "A Solution to the Positivity Problem in the State Space Approach to Modelling Vector-Valued Times Series," Working Paper, Kingston: Univ. of Rhode Island, 1991.

White, H., Asymptotic Theory for Econometricians, NY: Academic Press, 1984.